January, 1995

LIDS- P 2281

Research Supported By:

ARO grant DAAL03-92-G-0115 (CICS)
AFOSR grant F49620-92-J-0002
NSF grant MIP-9015281

Robust Multiscale Representation of Processes and Optimal Signal Reconstruction

Krim, H.
Pesquet, J.-C.
Willsky, A.S.

LIDS-P-2281
January 1995

ROBUST MULTISCALE REPRESENTATION OF PROCESSES AND OPTIMAL SIGNAL RECONSTRUCTION

H. Krim, J.-C. Pesquet‡, A. S. Willsky

Stochastic Systems Group, LIDS, MIT,
Cambridge MA 02139
‡Laboratoire des Signaux et Systèmes,
CNRS/UPS and GDR TdSI,
ESE, 91192 Gif sur Yvette Cédex, France

ABSTRACT

We propose a statistical approach to obtain a "best basis" representation of an observed random process. We derive statistical properties of a criterion first proposed to determine the best wavelet packet basis, and, proceed to use it in constructing a statistically sound algorithm. For signal enhancement, this best basis algorithm is followed by a nonlinear filter based on the Minimum Description Length (MDL) criterion. We show that it is equivalent to a *Min-Max* based algorithm proposed by Donoho and Johnstone.

1. Introduction

Wavelet theory and applications have been the focus of research over the last few years [1, 2]. The applications in a stochastic setting, however, have only recently been considered [3, 4, 5, 6].

Wavelet/wavelet packet based optimal representations have, for the most part, been carried out in a deterministic setting. In [3], a Karhunen-Loève approximation was obtained with the assumption that the wavelet coefficients remained uncorrelated.

A similar problem was investigated in [6, 7] for signal reconstruction. The goal was to filter the noise and obtain the best estimate of the signal. The adaptivity of the basis to a given process was, however, not addressed.

In this paper, we address the problem of signal reconstruction by using a layered optimization/filtering approach. The optimization step aims at unraveling a best basis of a process, by using its statistics. To reduce the effect of noise, the "optimal" representation of a process is then followed by a nonlinear filter which we derive using the Minimum Description Length principle. The latter criterion, as will be shown, turns out to be equivalent to that in [6] and mentioned in [8].

The paper is organized as follows: the next section briefly describes the background material required for the remainder of the paper. In Section 3, we develop a statistical best basis search algorithm for an observed random process. In Section 4, we derive a nonlinear filter for enhancing the signal representation, assuming that the noise is white or had been whitened, and we conclude with some remarks in Section 5.

The work of the first and third authors is partially supported by grants from ARO (DAAL03-92-G-0115) (Center for Intellingent Control), AFOSR (F49620-92-J-0002), and NSF (MIP-9015281).

2. Background

The nonoptimality of a wavelet for every process is overcome by making the multiscale analysis adaptive via wavelet packets.

A wavelet packet decomposition [9] is an extension of a wavelet representation, and allows a best matched analysis to a signal. To define wavelet packets, we first need to introduce functions of $L^2(\mathcal{R})$, $W_m(t)$, $m \in \mathcal{N}$, such that

$$\int_{-\infty}^{\infty} W_0(t) = 1, \tag{1}$$

and, for all $(k,j) \in \mathcal{Z}^2$,

$$\begin{aligned} 2^{-\frac{1}{2}} W_{2m}(\frac{t}{2}-k) &= \textstyle\sum_{l=-\infty}^{\infty} h_{l-2k}\, W_m(t-l), \\ 2^{-\frac{1}{2}} W_{2m+1}(\frac{t}{2}-k) &= \textstyle\sum_{l=-\infty}^{\infty} g_{l-2k}\, W_m(t-l), \end{aligned}$$

where $\{h_k\}_{k\in\mathcal{Z}}$ and $\{g_k\}_{k\in\mathcal{Z}}$ are the previously defined impulse responses of the QMF filters. If, for every $j \in \mathcal{Z}$, we define the vector space

$$\Omega_{j,m} \triangleq \mathrm{Span}\{W_m(t/2^j - k), k \in \mathcal{Z}\},$$

it can then be shown [9] that

$$\mathrm{L}^2(R) = \overset{\perp}{\bigoplus}_{(j.m)/I_{j,m}\in\mathcal{P}} \Omega_{j,m}, \tag{2}$$

for every partition $\mathcal{P}$ of $\mathcal{R}^+$ in intervals $I_{j,m} = [m2^{-j}, (m+1)2^{-j}[$, (i.e. $\{2^{-j/2}W_m(2^{-j}t-k), k \in \mathcal{Z}, (j,m)/I_{j,m} \in \mathcal{P}\}$) is an orthonormal basis of $\mathrm{L}^2(\mathcal{R})$. Such a basis is called a wavelet packet. The coefficients resulting from the decomposition of a signal $\mathrm{x}(t)$ in this basis are

$$C_{j,m}^k(\mathrm{x}) \triangleq < \mathrm{x}(t), \frac{1}{2^{j/2}} W_m(\frac{t}{2^j} - k) >, \tag{3}$$

with j denoting the resolution level, m the spectral bin and k the translation parameter.

3. Robust Best Basis

As noted earlier, a deterministic measure of energy concentration (MEC) has thus far been used in searching for a best basis, even in noisy signal scenarios. The noise, however, introduces a high variability in the resulting bases and makes this approach hard to justify. Our goal, here, is to use statistics from the observed process to achieve its best representation and to enhance its robustness to noise.

We assume the following signal model with an additive zero mean Gaussian white noise $\mathrm{b}(t)$ of power spectral density σ^2. [1]

$$\mathrm{x}(t) = \eta(t) + \mathrm{b}(t), \tag{4}$$

The signal $\eta(t)$ is assumed unknown and the observation of the process is carried out over $[0,T]$, where T is adequately selected for the observed process.

The wavelet packet coefficients $\{C_{j,m}^k(\mathrm{x}),\ k \in \mathcal{Z}, (j,m) \in \mathcal{N}^2\}$ are also normal with means $C_{j,m}^k(\eta)$ and variances $\mathrm{E}\{\mid C_{j,m}^k(\mathrm{b}) \mid^2\}$.

3.1. Analysis of Variance

In analyzing the forementioned process, two cases arize:

a. there is only noise in some subbands of the process ;

b. there is signal and noise.[2]

It is clear that we would ideally like to differentiate the cases "a" and "b" by using an objective adapted measure. To proceed, we first state the following straightforward result:

[1] If the psd is unknown, it must be estimated.

[2] The signal only case does not arize by assumption.

Lemma 1 *The power of the sequence $\{C_{j,m}^k(\mathrm{x})\}_{0\leq k<K2^{-j}}$ of $\mathrm{x}(t)$ at scale j, is*

$$\begin{aligned}\sigma_{j,m}^2 &= \frac{1}{K2^{-j}} \sum_{k=0}^{K2^{-j}-1} \mathrm{E}\{\mid C_{j,m}^k(\mathrm{x}) \mid^2\} \\ &= \frac{1}{K2^{-j}} \sum_{k=0}^{K2^{-j}-1} \mid C_{j,m}^k(\eta) \mid^2 + \sigma^2.\end{aligned}$$

As a consequence, we find in case "a", that $\sigma_{j,m}^2 = \sigma^2$, whereas it will generally be j and m dependent, in case "b". The scale invariance of the variance, which is unique to white noise, may then be part of a test in the search of an optimal basis. The null hypothesis will consist of the equality of the powers of two children born out of the same node of the tree, or

$$\begin{aligned}\mathrm{H}_0 : \sigma_{j,2m}^2 &= \sigma_{j,2m+1}^2 \\ \mathrm{H}_1 : \sigma_{j,2m}^2 &\neq \sigma_{j,2m+1}^2.\end{aligned}$$

Case "a" can be easily characterized (i.e. the variables are χ^2 with $2^{-j}K$ degrees of freedom at scale j), and the comparison can be based on an F-distribution test [10]. This test by itself, however, is not sufficiently robust, as ambiguities may arize at some nodes of the analysis. Additional robustness can be achieved by introducing redundancy in the test for the *noise-only* case (or its absence) and is dicussed next.

3.2. Statistical MEC

An entropy-like criterion was proposed as an efficient MEC [9]. The lower the measure, the more concentrated the coefficients are, the more adapted the basis is to the signal. This cost function, however, is a random variable, in the presence of noise, and should be viewed as such. Following Wickerhauser's rationale, we carry out the search using the MEC. Note that the MEC is now a random variable, thus any comparison needs to be statistically carried out. To proceed, properties of the MEC

$$\mathcal{I}(\{C_{j,m}^k\}) = \sum_{k=0}^{K2^{-j}-1} \mid C_{j,m}^k \mid^2 \log(\mid C_{j,m}^k \mid^2), \tag{5}$$

are first given below:

Proposition 1 *If $\{C_{j,m}^k\}_{0\leq k<K2^{-j}}$ is an i.i.d. sequence, then*

$$\frac{\mathcal{I}(\{C_{j,m}^k\}) - K2^{-j}\mu}{\sqrt{K}2^{-j/2}\epsilon} \sim \mathrm{N}(0,1), \quad K2^{-j} \to \infty, \tag{6}$$

where

$$\mu = \mathrm{E}\{C_{j,m}^k\}, \tag{7}$$

$$\epsilon^2 = \mathrm{Var}\{C_{j,m}^k\}. \tag{8}$$

Furthermore, when $\mathcal{C}_{j,m}^k$ is $N(0,\sigma^2)$, we have

$$\begin{aligned}\mu &= \sigma^2(2-\log 2-\gamma+2\log\sigma),\\ \epsilon^2 &= \sigma^4[2\log^2(2)-12\log 2+4+\frac{3\pi^2}{2}+\\ &\quad 4(\log 2-3)\gamma+2\gamma^2+8\log^2(\sigma)\\ &\quad +(24-8\log 2-8\gamma)\log\sigma], \qquad (9)\end{aligned}$$

where γ denotes the Euler's constant.[3]

Proof: This follows after some algebra and by later invoking the central limit theorem to prove the asymptotic normality of the MEC criterion $\mathcal{I}(\{\mathcal{C}_{j,m}^k\})$. More details can be found in [10] ■

Note that the data record length K should be sufficiently high at the given resolution j for the asymptotic behavior, previously derived to hold. The comparison of the MEC cost functions between children (i.e. $\mathcal{I}(\{\mathcal{C}_{j,2m+1}^k\})$ and $\mathcal{I}(\{\mathcal{C}_{j,2m}^k\})$ and the corresponding parent ($\mathcal{I}(\{\mathcal{C}_{j,m}^k\})$) can now be simply carried out via a hypothesis test. Note that this can take place only if the cross-covariance between each of the children's coefficients and the corresponding parent's is known. This is shown in detail in [10]. These properties will allow us to couple the MEC to the variance test to obtain a best basis. A brief summary of the algorithm is given next.

3.3. Algorithm

The details of the following summarized algorithm may be found in [10]:

Algorithm

Select an overall significance level of the test, α

$\forall m \in \{0,\dots,2^{j_m-1}\}$, $\mathcal{H}_{j_m,m} \triangleq \mathcal{I}(\{\mathcal{C}_{j_m,m}^k\})$

$j = j_m - 1$

Begin

$\forall m \in \{0,\dots,2^j-1\}$,

1. Test

$H0: \ \sigma_{j+1,2m}^2 = \sigma_{j+1,2m+1}^2$

$H1: \ \sigma_{j+1,2m}^2 \neq \sigma_{j+1,2m+1}^2$

2. if $H0$ accepted [4]

Test

$$\begin{aligned}H2:\mathcal{H}_{j+1,2m} &+ \mathcal{H}_{j+1,2m+1} < \mathcal{I}(\{\mathcal{C}_{j,m}^k\})\\ H3:\mathcal{H}_{j+1,2m} &+ \mathcal{H}_{j+1,2m+1} \geq \mathcal{I}(\{\mathcal{C}_{j,m}^k\})\end{aligned}$$

If $H2$ accepted, repeat steps 1 and 2 with a significance level $\alpha_1 > \alpha$,

goto 4

3. Else test[5]

$$\begin{aligned}H2:\mathcal{H}_{j+1,2m} &+ \mathcal{H}_{j+1,2m+1} < \mathcal{I}(\{\mathcal{C}_{j,m}^k\})\\ H3:\mathcal{H}_{j+1,2m} &+ \mathcal{H}_{j+1,2m+1} \geq \mathcal{I}(\{\mathcal{C}_{j,m}^k\})\end{aligned}$$

4. If $H2$

$$\begin{aligned}\mathcal{P}_{j,m} &\triangleq \mathcal{P}_{j+1,2m}\cup\mathcal{P}_{j+1,2m+1}\\ \mathcal{H}_{j,m} &\triangleq \mathcal{H}_{j+1,2m}+\mathcal{H}_{j+1,2m+1}\end{aligned}$$

Else,

$$\begin{aligned}\mathcal{P}_{j,m} &\triangleq \{I_{j,m}\}\\ \mathcal{H}_{j,m} &\triangleq \mathcal{I}(\{\mathcal{C}_{j,m}^k\})\end{aligned}$$

$j = j-1$, if $j \geq 0$ goto begin

End

The statistically optimal partition results as $\mathcal{P}_{0,0}$.

4. Nonlinear Filtering

We assume an analysis of a signal $\mathrm{x}(t)$ on an interval whose reconstruction is achieved in an orthonormal wavelet packet basis:[6] By reindexing the coefficients with a single indexing subscript, we can reformulate the problem as one of estimating coefficients $\{\eta_n\}_{1\leq n\leq K}$ embedded in an additive $\mathrm{N}(0,\sigma^2)$ white noise, from observations $\{\mathrm{y}_n\}_{1\leq n\leq K}$. Since $\{\mathrm{y}_n\}_{1\leq n\leq K}$ represents the selected coefficients $\mathcal{C}_{\cdot,\cdot}^k(\mathrm{x}), k=\{1,\cdots,K\}$ of $\mathrm{x}(t)$ resulting from the best basis search, we would expect the latter to be adequately represented by a small number P of orthogonal directions, in contrast to noise, which necessarily would be present in all the directions being considered.

To proceed, we first assume, that we have reordered, if need be, the coefficients $\{\mathrm{y}_n\}_{1\leq n\leq K}$ such that

$$|\,\mathrm{y}_1\,| \geq |\,\mathrm{y}_2\,| \geq \cdots \geq |\,\mathrm{y}_K\,|, \qquad (10)$$

[3]The Euler's constant is defined by $\gamma = \lim_{n\to\infty}\sum_{k=1}^n \frac{1}{k} - \log n \simeq 0.5772$.

[4]Note that the random variables are central χ^2 variables.

[5]The random variables are now noncentral and are a function of η

[6]This approach may also be applied to wavelet decomposition, which appear as a special case.

and appropriately reindexed them. The independence is assumed among the coefficients, and their joint probability density function straightforwardly results as,

$$f(y_1,\ldots,y_K \mid \zeta) = \frac{1}{(2\pi)^{K/2}} \times e^{-\frac{1}{2\sigma^2}\left(\sum_{l=1}^{P}(y_{n_l}-\eta_{n_l})^2+\sum_{l=P+1}^{K}y_{n_l}^2\right)},$$

where $\zeta = (n_1,\ldots,n_P,\eta_{n_1},\ldots,\eta_{n_P})^T$. With the density in hand, we call upon the MDL criterion [11], to optimally encode the coefficients by finding the shortest code length which summarizes the data. The joint probability density function, is in turn used to construct the MDL functional whose minimization leads to the following result,

Proposition 2 *The P coefficients which, based upon the MDL criterion, give the optimal coding length, are determined by the components which satisfy the following inequality:*

$$\mid y_n \mid > \sigma\sqrt{2\log K}. \tag{11}$$

The details of the above result can be found in [10]. ■ This is intuitively appealing as the signal enhancement, in a sense, results from picking the signal components (or "signal subspace") and eliminating the noise components (or "noise subspace") by a nonlinear procedure. This is precisely the "threshold" that Donoho and Johnstone [6] derived from a *Min-Max* perspective.

5. Conclusion

Assuming the noise white and gaussian in our signal model, we studied a criterion which was successfully used for deterministic optimal signal representation by way of wavelet packets, and derived its statistical properties. These properties were coupled with an intrinsic invariance of white noise variance under such an orthonormal transformation, to construct an algorithm which objectively and systematically resulted in a statistically best basis. The *best basis* helped a MDL-based nonlinear filter in further "pruning" redundant components, and resulted in an enhanced signal reconstruction.

REFERENCES

[1] S. Mallat. "A theory for multiresolution signal decomposition: the wavelet representation". *IEEE Trans. on Pattern Anal. and Mach. Intell.*, PAMI-11:674–693, Jul. 1989.

[2] I. Daubechies. "Orthonormal bases of compactly supported wavelets". *Com. Pure and Appl. Math.*, XLI:909–996, 1988.

[3] G. W. Wornell. "A Karhunen-Loève-like expansion for $1/f$ processes via wavelets". *IEEE Trans. Inform. Theory*, 36:859–861, Jul. 1990.

[4] P. Flandrin. "Wavelet analysis and synthesis of fractional Brownian motion". *IEEE Trans. Inform. Theory*, IT-38:910–917, Mar. 1992.

[5] M. Basseville, M. Benveniste, A. Chou, K.C. Golden, R. Nikoukhah, and A. S. Willsky. "Modeling and Estimation of Multiresolution Stochastic Processes". *IEEE Trans. Inform. Theory*, IT-38:529–532, Mar. 1992.

[6] D. L. Donoho and I. M Johnstone. "Ideal spatial adaptation by wavelet shrinkage". preprint Dept. Stat., Stanford Univ., Jun. 1992.

[7] B. Lumeau, J.-C. Pesquet, J.-F. Bercher, and L. Louveau. "Optimisation of bias-variance trade-off in non parametric spectral analysis by decomposition into wavelet packet". In Y. Meyer and S. Roques, editors, *Progress in wavelet analysis and applications*, pages 285–290. Editions Frontieres, Gif sur Yvette, 1993.

[8] P. Moulin. "A wavelet regularization method for diffuse radar target imaging and speckle noise reduction". *Journ. Math. Imaging and Vision*, 3:123–134, 1993.

[9] M. V. Wickerhauser. "INRIA lectures on wavelet packet algorithms". In *Ondelettes et paquets d'ondelettes*, pages 31–99, Roquencourt, Jun. 17-21 1991.

[10] H. Krim and J.-C. Pesquet. "Robust Multiscale Representation of Processes and Optimal Signal Reconstruction". in preparation.

[11] J. Rissanen. "Modeling by Shortest Data Description". *Automatica*, 14:465–471, 1978.